



TITLE:

最適化の手法 (非線型方程式の数値解析)

AUTHOR(S):

柳井, 浩

CITATION:

柳井, 浩. 最適化の手法 (非線型方程式の数値解析). 数理解析研究所講究録 1973, 190: 71-96

ISSUE DATE:

1973-10

URL:

<http://hdl.handle.net/2433/107236>

RIGHT:

最適化の手法

慶大 エ学部 柳井 浩

§ 1 序

この研究は制約条件なしの最適化問題の解法に於いてよく用いられる conjugate direction method 系の諸手法の体系化の一つの試みである。conjugate direction method はもともと線形連立方程式の一解法として考案されたものであり、その体系化の試みも 2~3 あるようである。こゝでは、できるだけ発見的なやり方で、幾何学的な意味づけを失わぬよう Gram-Schmidt の直交化法からみちびかれる 2 つのタイプの共役化過程を中心的視角として conjugate direction method 系の諸手法を見わたすことにした。

本稿の構成は、まず、§ 2 に於いて我々のとりあつかう目的関数 $f(x) = \frac{1}{2}x^T A x + b^T x + C$ の 2~3 の性質と；直線上に於ける最小化の反復としての最小化過程；そのうちでも緩和法は、 A が対角型の場合、変数の数だけの回数の直線上の最小化の反復で最小点をもたらす事等基本的な事実が示されている。

次に，§3に於いては， A を対角型になおすための共役基底の導入とその構成法， A ， A' および E の共役基底による表現，それを与える漸化式を示した。

共役基底の構成法のうちでも，本稿で主としてあつかう，Gram-Schmidtの直交化にもとづくものは節をあらため§4でのべた。Gram-Schmidtの直交化にもとづく共役化の方法として，2つのタイプのものが我々の目的に役立ちそうに思われる。Type 1は直交化のうち適当な座標変換を行なうものであり，Type 2はそれまでに共役化されたベクトルに A を乗じたものと直交するベクトルを求めるものである。

この節ではさらに，これらの過程をベクトルに関する漸化式として書き，その形をもつ一連の方法を例としてあげた。

§5では，§4の漸化式をさらに行列型に書きあらためると共に，§3に於いて与えた A' を得る漸化式と組み合わせ，最小化と同時に A' が得られるアルゴリズムを論じ，その形をもつ一連の方法を例としてあげた。

§2以下では，文章による説明は少ししか与えていないが，表題，式，図等で充分なように配慮したつもりである。

§ 2. Preliminaries

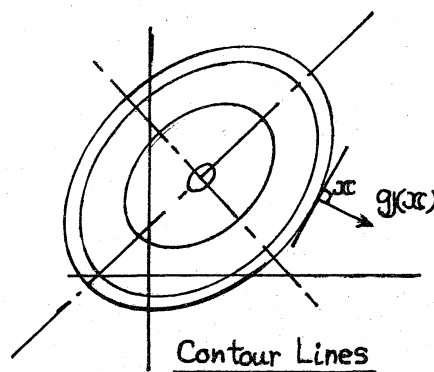
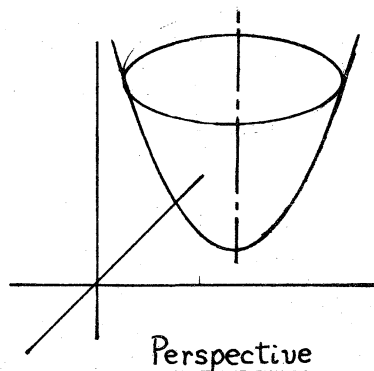
◎我々の目的函数, その勾配, 最小点, 等に関する性質は次の通りである。

2.1 Objective Function

$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in R^n, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in R^n, \quad c: \text{real}$$

$n \times n$ A : real, symmetric, positive definite



2.2 Gradients

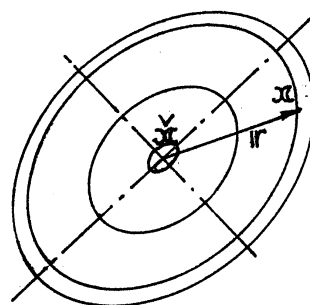
$$g(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}, \quad g \cdot \equiv g(x \cdot), \quad g(x) = Ax + b$$

2.3 Minimum Point

$$f(x): \text{Min at } \check{x} \Leftrightarrow g(\check{x}) = 0 \Leftrightarrow A\check{x} + b = 0 \Leftrightarrow \check{x} = -A^{-1}b$$

2.4 Residual $r = x - \tilde{x} = A^{-1}g(x)$

$$\begin{aligned} r &= x - \tilde{x} = x - (-A^{-1}b) = \\ &= A^{-1}Ax + A^{-1}b \\ &= A^{-1}(Ax + b) = A^{-1}g(x) \end{aligned}$$



◎本稿では，直線上の最小化を单元とする最小化過程を考える。これをめぐる諸式は次の通りである。

2.5 Linear Minimization

Straight Line $x = x_m + \xi t$

base point

search direction

Objective Function on $x = x_m + \xi t$

$$\varphi(\xi) \equiv f(x_m + \xi t)$$

$$\varphi'(\xi) = t^T g(x_m + \xi t)$$

$$= t^T (Ax_m + b) + \xi t^T A t$$

$$= t^T g_m + \xi t^T A t$$

Min. Point on $x = x_m + \xi t$

$$x_{m+1} = x_m + \xi_m t$$

$$\underbrace{\quad}_{\text{min. point}} \quad \underbrace{\xi_m}_{\xi_m = \mu(x_m, t)}$$

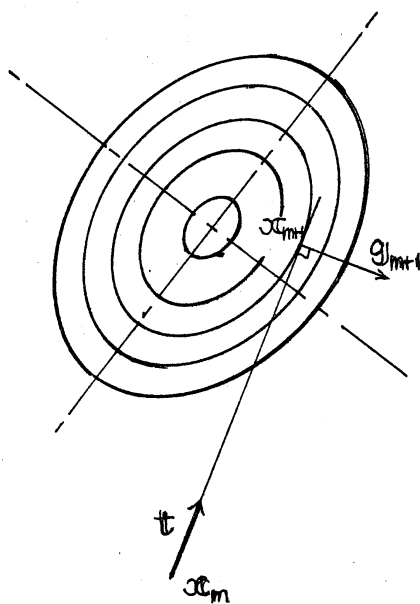
$$\underline{2.6} \quad \mu(x_m, t) = -\frac{t^T g_m}{t^T A t} \quad \leftarrow \varphi'(\xi_m) = 0$$

Lemma 1

2.7

$$t^T g_{m+1} = 0$$

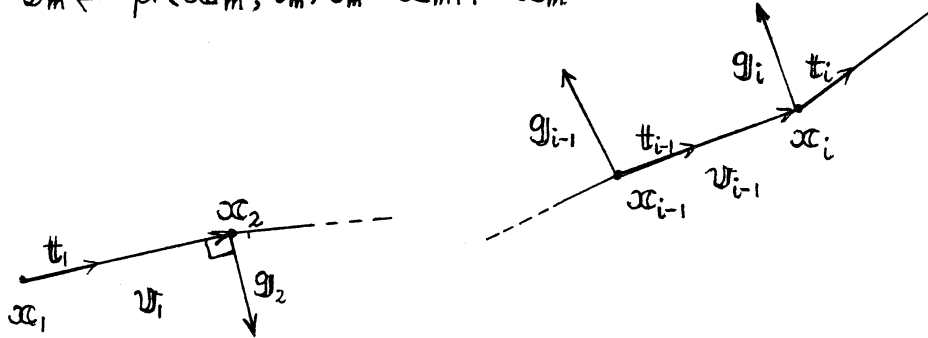
$$\therefore 0 = \varphi'(\xi_m) = t^T g(x_m + \xi_m t) = t^T g_{m+1}$$



2.8 Minimization Process

$$x_{m+1} = x_m + w_m$$

$$w_m \equiv \mu(x_m, t_m) t_m = x_{m+1} - x_m$$



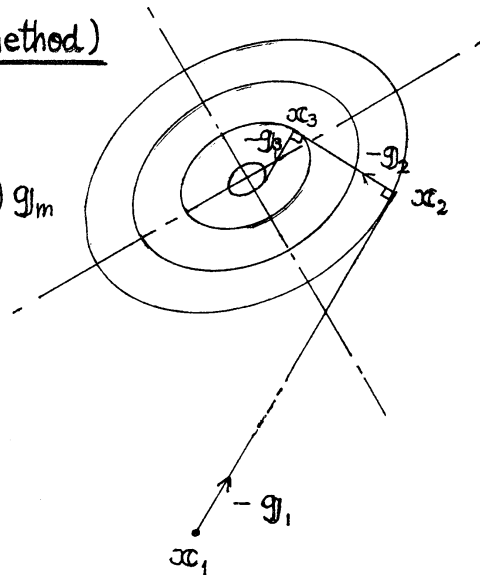
$$y_{i-1} \equiv g_i - g_{i-1} = A w_{i-1}$$

$$\begin{aligned} \therefore g_i - g_{i-1} &= A x_i + b - (A x_{i-1} + b) = \\ &= A(x_i - x_{i-1}) = \\ &= A w_{i-1} \end{aligned}$$

2.9 Example (Optimal Gradient Method)

$$t_m = -g_m$$

$$x_{m+1} = x_m - \mu(x_m, -g_m) g_m$$



◎ 最小化過程の最も素朴な例は，次の緩和法であるが，これは関数の形 — いゝかえれば座標のえらび方で収束性が大巾に変わる。

2.10 Relaxation Method

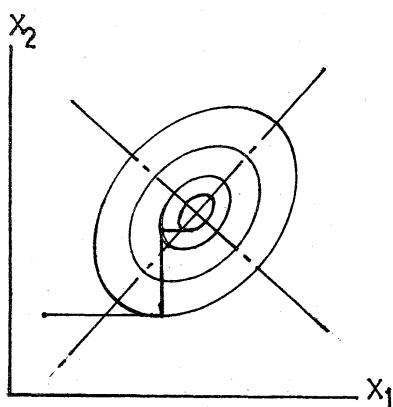
$$t_1 = e_1, t_2 = e_2, \dots, t_n = e_n$$

$$t_{n+1} = e_1, t_{n+2} = e_2, \dots, t_{2n} = e_n$$

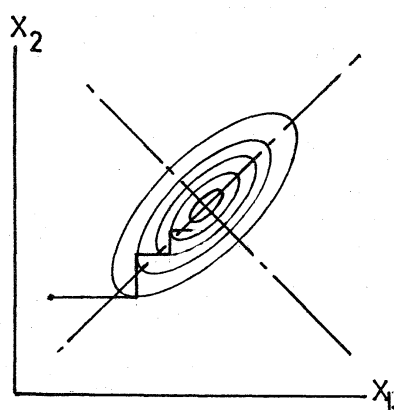
...

$$x_{m+1} = x_m + \mu(x_m, e.) e.$$

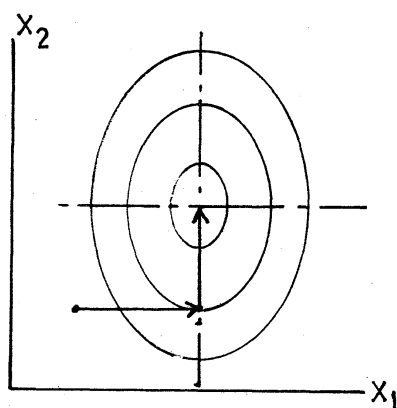
$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Bigg\} i$$



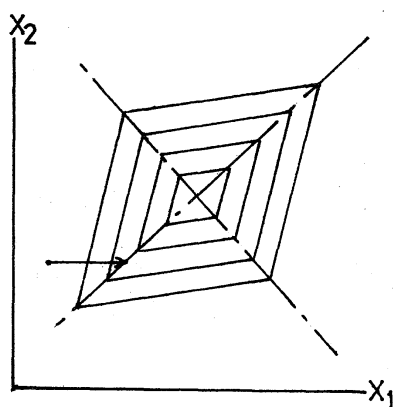
a



b



c



d

◎ 上図から明らかなように (C) の場合が最も都合がよい。これは式でいえば次の通りである。

2.11

Lemma 2

$$A = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \\ & & & \lambda_n \end{pmatrix} : \begin{array}{l} \text{diagonal,} \\ \text{positive definite} \\ (\lambda_1, \lambda_2, \dots, \lambda_n > 0) \end{array}$$

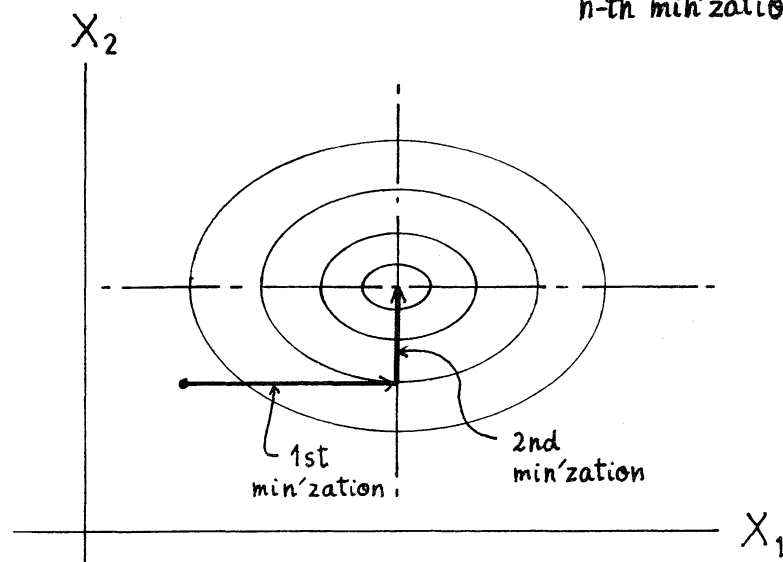
$$\Downarrow$$

By relaxation method, x_m reaches \check{x}
at most in n linear minimizations, i.e. $m \leq n+1$

Proof

$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

$$= \underbrace{\left(\frac{1}{2} \lambda_1 x_1^2 + b_1 x_1 \right)}_{\substack{\downarrow \\ \min \\ \text{in} \\ \text{1st min'zation}}} + \underbrace{\left(\frac{1}{2} \lambda_2 x_2^2 + b_2 x_2 \right)}_{\substack{\downarrow \\ \min \\ \text{in} \\ \text{2nd min'zation}}} + \dots + \underbrace{\left(\frac{1}{2} \lambda_n x_n^2 + b_n x_n \right)}_{\substack{\downarrow \\ \min \\ \text{in} \\ \text{n-th min'zation}}} + c$$



◎そこで、座標変換によって、 A を対角化することが考えられる。2.12はそのための準備。

2.12 Coordinate Transformations OLD

	COORDINATE	new COORDINATE
BASIS :	e_1, e_2, \dots, e_n	h_1, h_2, \dots, h_n
P :	$P = x_1 e_1 + x_2 e_2 + \dots + x_n e_n$	$P = z_1 h_1 + z_2 h_2 + \dots + z_n h_n$
COORDINATE :	$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$	$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$
VECTORS :		

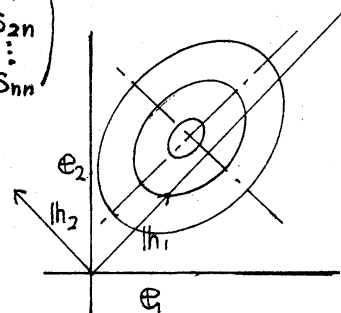
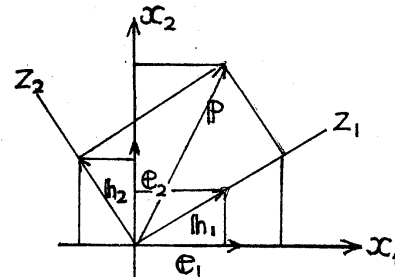
RELATIONS:

$$\begin{aligned} h_1 &= s_{11} e_1 + s_{21} e_2 + \dots + s_{n1} e_n \\ h_2 &= s_{12} e_1 + s_{22} e_2 + \dots + s_{n2} e_n \\ &\vdots \\ h_n &= s_{1n} e_1 + s_{2n} e_2 + \dots + s_{nn} e_n \end{aligned}$$

$$S_1 = \begin{pmatrix} s_{11} \\ s_{21} \\ \vdots \\ s_{n1} \end{pmatrix}, \quad S_2 = \begin{pmatrix} s_{12} \\ s_{22} \\ \vdots \\ s_{n2} \end{pmatrix}, \quad \dots, \quad S_n = \begin{pmatrix} s_{1n} \\ s_{2n} \\ \vdots \\ s_{nn} \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \vdots & \vdots & & \vdots \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$$

$$x = S z, \quad S = (S_1, S_2, \dots, S_n)$$

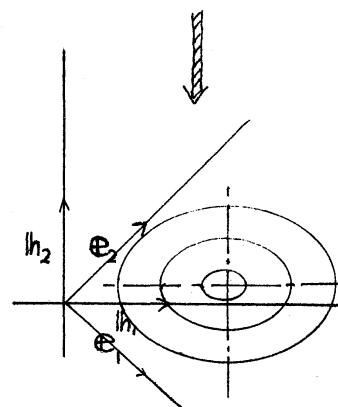


2.13 $f(x)$ in terms of z

$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

$$f_1(z) = \frac{1}{2} z^T S^T A S z + (S^T b)^T z + c$$

$$x = S z$$



◎ A の対角化のためには, 互に共役なベクトルからなる基底, 共役基底をとればよい。 - - -

§ 3 Conjugate Directions

3.1 Conjugate Relations

$$\boxed{\begin{array}{l} p : \text{conjugate to } q \\ \text{with respect to } A \\ p, q \in \mathbb{R}^n \\ \neq 0 \end{array}} \equiv \boxed{p^T A q = 0}$$

3.2 Canonical Basis

$$\boxed{\{p_1, p_2, \dots, p_n\} : (A)\text{-canonical basis}} \equiv \boxed{\begin{array}{l} \{p_1, p_2, \dots, p_n\} : \text{basis} \\ p_i \perp p_j \text{ conjugate} \\ i \neq j \end{array}}$$

3.3 Reduction of Quadratic Forms to Canonical Forms

$\{p_1, p_2, \dots, p_n\}$: canonical basis in terms of a given basis

Coordinate Transformation

$$\begin{array}{ccc} \text{OLD} & & \text{NEW} \\ \downarrow & & \downarrow \\ x & = & Pz, \quad P \equiv (p_1 | p_2 | \dots | p_n) \end{array}$$

$$\begin{array}{ccc} \text{OLD} & & \text{NEW} \\ \hline f(x) = \frac{1}{2} x^T A x + b^T x + c & \Bigg| & f_1(z) = \frac{1}{2} z^T P^T A P z + (P^T b)^T z + c \end{array}$$

$$\begin{aligned} P^T A P &= \begin{pmatrix} p_1^T \\ p_2^T \\ \vdots \\ p_n^T \end{pmatrix} A (p_1 | p_2 | \dots | p_n) \\ &= \begin{pmatrix} p_1^T A p_1 & p_1^T A p_2 & \dots & p_1^T A p_n \\ p_2^T A p_1 & p_2^T A p_2 & \dots & p_2^T A p_n \\ \vdots & \vdots & \ddots & \vdots \\ p_n^T A p_1 & p_n^T A p_2 & \dots & p_n^T A p_n \end{pmatrix} \end{aligned}$$

$p_i^T A p_j = 0 \quad i \neq j$

: Canonical Form
(Diagonal)

* $p_i^T A p_i$: canonical coefficient $i=1, 2, \dots, n$

◎ 共役基底をつかえば, A, A^{-1}, E を次のように表現することができる。また, これらを漸化式の形で書いておくと, 最小点と同時に A^{-1} 等を計算するアルゴリズムが構成できる。

3.4 Representation of A, A^{-1} and E in Canonical Basis

a. Lemma 3

$$\boxed{\{p_1, p_2, \dots, p_n\}: \text{canonical basis}} \Rightarrow \boxed{A = \sum_{j=1}^n \frac{A p_j p_j^T A}{p_j^T A p_j}}$$

Proof 1

* $\forall x \in \mathbb{R}^n \exists \{x_1, x_2, \dots, x_n\}$ such that

$$x = x_1 p_1 + x_2 p_2 + \dots + x_n p_n$$

$$* Ax = x_1 A p_1 + x_2 A p_2 + \dots + x_n A p_n = \sum_{j=1}^n x_j A p_j$$

$$* \sum_{j=1}^n \frac{A p_j p_j^T A}{p_j^T A p_j} x = \sum_{j=1}^n \frac{A p_j p_j^T A}{p_j^T A p_j} \sum_{k=1}^n x_k p_k = \sum_{j=1}^n x_j A p_j = Ax$$

Proof 2 (Heuristics)

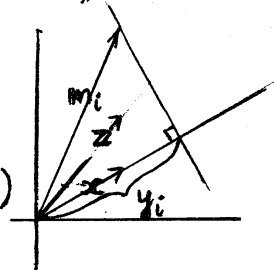
$$(i) \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ given}$$

$$\|x\| = 1 \quad (\text{without loss of generality})$$

$$y = M x, \quad M = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{pmatrix}$$

$$m_i = \frac{y_i}{z^T x} z^T \quad z^T x \neq 0$$

$$M = \frac{y z^T}{z^T x} \quad \|z\| = 1 \quad (\text{without loss of generality})$$



$$(ii) \quad \frac{A_i p_i}{y} = \frac{A_i p_i}{x} \quad z = A p_i$$

$$A_i = \frac{A p_i p_i^T A}{p_i^T A p_i} \quad \text{no effects on } [A p_i]^\perp$$

$$\therefore A = \sum_{i=1}^n \frac{A p_i p_i^T A}{p_i^T A p_i}$$

b. Corollary 3.1

$$\boxed{A_i = \sum_{j=1}^i \frac{A P_j P_j^T A}{P_j^T A P_j}} \Rightarrow \boxed{A P_i = A_i P_i}$$

c. Corollary 3.2

$$\boxed{\begin{aligned} A_1 &= \frac{A P_1 P_1^T A}{P_1^T A P_1} \\ A_i &= A_{i-1} + \frac{A P_i P_i^T A}{P_i^T A P_i} \end{aligned}} \Rightarrow \boxed{A_n = A}$$

d. Corollary 3.3

$$\boxed{y_i = A P_i} \Rightarrow \boxed{A^{-1} = \sum_{j=1}^n \frac{A^{-1} y_j y_j^T A^{-1}}{y_j^T A^{-1} y_j} = \sum_{j=1}^n \frac{P_j P_j^T}{P_j^T A P_j}}$$

$$\therefore y_i^T A^{-1} y_j = P_i^T A A^{-1} A P_j = P_i^T A P_j = 0$$

e. Corollary 3.4

$$\boxed{\begin{aligned} E &= \sum_{j=1}^n \frac{P_j P_j^T A}{P_j^T A P_j} = \sum_{j=1}^n \frac{A P_j P_j^T}{P_j^T A P_j} \\ &= \sum_{j=1}^n \frac{y_j y_j^T A^{-1}}{y_j^T A^{-1} y_j} = \sum_{j=1}^n \frac{A^{-1} y_j y_j^T}{y_j^T A^{-1} y_j} \end{aligned}}$$

E: unit matrix

◎そこで，互に共役な方向に直線最小化をくりかえせば，高々 n 回で最小点に到達する。

3.5 Minimization along Conjugate Directions

a. Conjugate Directions Methods

$$\begin{aligned} x_{m+1} &= x_m + v_m \\ v_m &= \mu(x_m, t_m) t_m \\ t_1, t_2, \dots, t_n &: \text{canonical basis} \\ (v_i^T A v_j &= 0 \quad i \neq j) \end{aligned}$$

b. Exactness of Conjugate Directions Methods

By Conjugate Directions Methods, x_m reaches \check{x} at most in n linear minimizations, i.e. $m \leq n+1$

Proof See Lemma 2. and 3.3 ■

◎次の補助定理は共役方向に進むときに見られる一般的性質。

c. Lemma 4

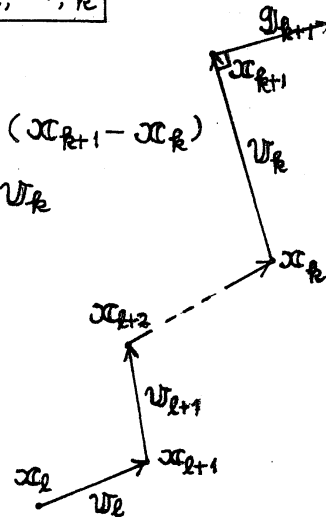
$$g_{k+1}^T v_\ell = 0 \quad \begin{array}{l} k = 1, 2, \dots, n \\ \ell = 1, 2, \dots, k \end{array}$$

Proof.

$$\begin{aligned} x_{k+1} &= x_{l+1} + (x_{l+2} - x_{l+1}) + \dots + (x_{k+1} - x_k) \\ &= x_{l+1} + v_{l+1} + v_{l+2} + \dots + v_k \\ &= x_{l+1} + \sum_{i=l+1}^k v_i \end{aligned}$$

$$\begin{aligned} g_{k+1} &= A x_{k+1} + b \\ &= A x_{l+1} + b + \sum_{i=l+1}^k A v_i \\ &= g_{l+1} + \sum_{i=l+1}^k A v_i \end{aligned}$$

$$g_{k+1}^T v_\ell = \underbrace{g_{l+1}^T v_\ell}_0 \text{ (Lemma 1)} + \sum_{i=l+1}^k \underbrace{v_i^T A v_\ell}_0 \text{ (a)} = 0$$



◎ 共役基底を構成するにはいくつかの方法がある。

3.6 Construction of Canonical Basis

a. Complete Squares

$$\begin{aligned}
 \text{Ex. } (x_1, x_2) \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
 &= a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 \\
 &= a_{11}x_1^2 - \frac{a_{12}^2}{a_{22}}x_1^2 + a_{22} \left(\frac{a_{12}^2}{a_{22}^2}x_1^2 + 2\frac{a_{12}}{a_{22}}x_1x_2 + x_2^2 \right) \\
 &= \frac{a_{11}a_{22} - a_{12}^2}{a_{22}}x_1^2 + a_{22} \left(\frac{a_{12}}{a_{22}}x_1 + x_2 \right)^2 \\
 &= \frac{a_{11}a_{22} - a_{12}^2}{a_{22}}Z_1^2 + a_{22}Z_2^2
 \end{aligned}$$

$$Z_1 \equiv x_1$$

$$Z_2 \equiv \frac{a_{12}}{a_{22}}x_1 + x_2$$

b. Jacobi's Method

$$\begin{aligned}
 \text{Ex. } \left\{ \begin{array}{l} p_1 = e_1 \\ p_2 = \alpha_1^1 e_1 + e_2 \\ p_3 = \alpha_1^2 e_1 + \alpha_2^2 e_2 + e_3 \end{array} \right. \quad e_1, e_2, e_3 : \text{given basis}
 \end{aligned}$$

$$p_2^T A e_1 = 0$$

$$\alpha_1^1 e_1^T A e_1 + e_2^T A e_2 = 0$$

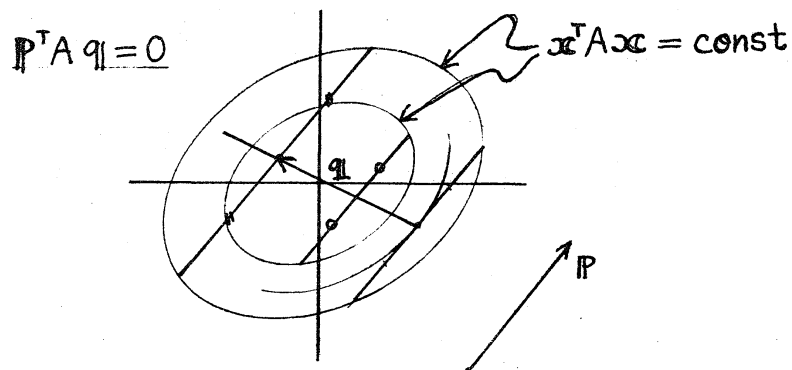
$$p_3^T A e_1 = 0$$

$$\alpha_1^2 e_1^T A e_1 + \alpha_2^2 e_2^T A e_1 + e_3^T A e_1 = 0$$

$$p_3^T A e_2 = 0$$

$$\alpha_1^2 e_1^T A e_2 + \alpha_2^2 e_2^T A e_2 + e_3^T A e_2 = 0$$

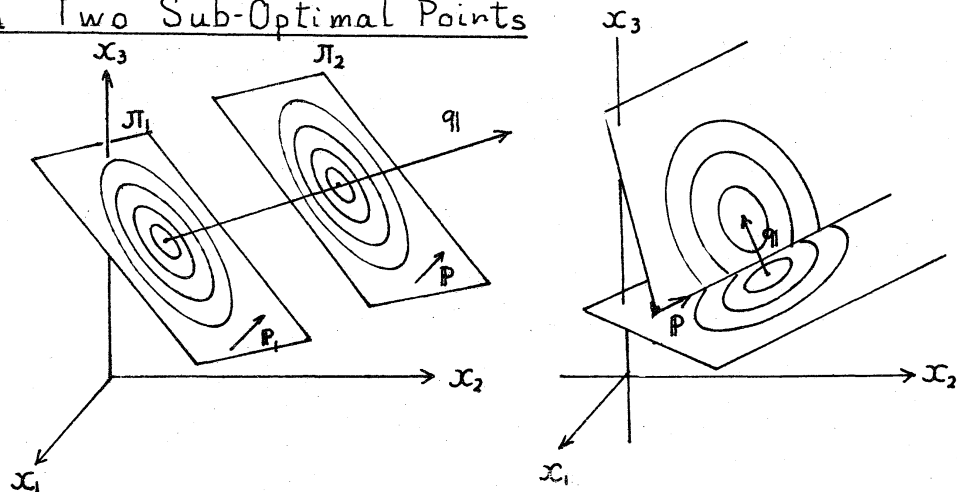
c. Middle Points of Chords



◎ 次の 2 つは，最適化によく用いられるものだが，本稿では

c を中心に考える。c については § 4 でくわしくのべる。

d. Two Sub-Optimal Points



$$P^T A q_1 = 0$$

cf. (PARallel TANgent Algorithm
Powell-Zangwill's Algorithm

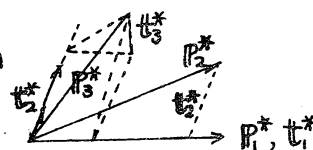
e. Gram Schmidt Orthogonalization Process

cf. (Fletcher Reeves's (Polak-Ribière) Algorithm
Davidon Fletcher Powell's Algorithm

◎ 以下では, Gram Schmidt 法にもとづく共役化を考える。

§4 Recurrence Relations in Vector Form

4.1 Gram Schmidt Orthogonalization



a. $p_1^*, p_2^*, \dots, p_n^*$ given basis

$$t_1^* \equiv p_1^*$$

$$t_i^* \equiv p_i^* - \sum_{j=1}^{i-1} \frac{p_i^{*T} t_j^*}{t_j^{*T} t_j^*} t_j^* \quad i=1, \dots, n$$

\Rightarrow

$$\{t_1^*, t_2^*, \dots, t_n^*\}:$$

orthogonal basis

◎ Type 1 は直交基底を座標変換によって共役基底になおす。

b. Construction of Canonical Basis — Type 1

(i) $* A$: positive definite $\Rightarrow \exists J$ such that

$$n \times n J: \text{regular}, J^T J = A$$

$* p, q$: conjugate $\Leftrightarrow p^* = Jp, q^* = Jq$: orthogonal

$$\therefore p^T A q = p^T J^T J q = (Jp)^T (Jq) = p^{*T} q^*$$

(ii) orthogonal basis canonical basis

$\{p_1^*, p_2^*, \dots, p_n^*\}$: given basis

$\{p_1, p_2, \dots, p_n\}$: given basis

$$p_i^* = J p_i$$

$$t_1^* \equiv p_1^*$$

$$t_i^* \equiv p_i^* - \sum_{j=1}^{i-1} \frac{p_i^{*T} t_j^*}{t_j^{*T} t_j^*} t_j^*$$

$$J t_1 = J p_1$$

$$J t_i = J p_i - \sum_{j=1}^{i-1} \frac{p_i^T J^T J t_j}{t_j^T J^T J t_j} J t_j$$

$J^{-1} \rightarrow$

$$t_1 = p_1$$

$$t_i = p_i - \sum_{j=1}^{i-1} \frac{p_i^T A t_j}{t_j^T A t_j} t_j$$

$\{t_1^*, t_2^*, \dots, t_n^*\}$ orthogonal basis

$\{t_1, t_2, \dots, t_n\}$ canonical basis

◎ Type 2 は, それまでに求めた共役系 $P_1 \cdots P_{i-1}$ に A を乗じた系 AP_1, \dots, AP_{i-1} と直交するベクトルを Gram-Schmidt 法で作る。

c. Construction of Canonical Basis — Type 2

$\{P_1, P_2, \dots, P_n\}$: given basis

$\begin{aligned} t_1 &= P_1 \\ u_1 &= At_1 \\ u_j &= At_j - \sum_{l=1}^{j-1} \frac{u_l^T At_j}{u_l^T u_l} u_l \\ t_i &= P_i - \sum_{j=1}^{i-1} \frac{P_i^T u_j}{u_j^T u_j} u_j \\ i &= 1, 2, \dots, n \end{aligned}$	$\left. \begin{array}{l} \\ \\ \\ \end{array} \right] \begin{array}{l} \text{orthogonal basis} \\ \text{for } [At_1, At_2, \dots, At_{j-1}] \\ \\ \text{vector orthogonal to} \\ [At_1, At_2, \dots, At_{i-1}] \end{array}$
--	---

$\{t_1, t_2, \dots, t_n\}$ canonical basis

◎ ここで, 最初に与える '任意の基底' として, 実際にかえそうなものをしらべておく。

4.2 Practical Selection of the "Given Basis"

a. $P_i \equiv e_i$

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

b. $p_i \equiv g_i$ (so far as $g_i \neq 0$)

(b' $p_i \equiv y_{i-1} = g_i - g_{i-1}$, $p_1 = g_1$)

Remark 1 $g_i \notin [t_1, t_2, \dots, t_{i-1}]$

$$\therefore \begin{cases} g_i \neq 0 \\ g_i^T t_j = 0, \quad j = 1, \dots, i-1 \end{cases}$$

Remark 2 $g_i \notin [At_1, At_2, \dots, At_{i-1}]$

$$\therefore (i) \quad g_i \in [At_1, At_2, \dots, At_{i-1}]$$

$$\Rightarrow \exists (\alpha_1, \alpha_2, \dots, \alpha_{i-1}) \text{ such that}$$

$$g_i = \alpha_1 At_1 + \dots + \alpha_{i-1} At_{i-1}$$

$$(ii) \quad t_j^T g_i = 0, \quad j = 1, 2, \dots, i-1$$

$$\therefore \begin{cases} t_j \parallel \psi_j \\ \psi_j^T g_i = 0 \quad j = 1, 2, \dots, j \end{cases} \quad (\text{Lemma 4})$$

$$(iii) \quad \alpha_j = 0 \quad j = 1, 2, \dots, i-1$$

$$\therefore \underset{(ii)}{0} = \underset{(i)}{t_j^T} g_i = \alpha_1 \underbrace{t_j^T A t_1}_{\parallel 0} + \dots + \alpha_j \underbrace{t_j^T A t_j}_{\neq 0} + \dots + \alpha_{i-1} \underbrace{t_j^T A t_{i-1}}_{\parallel 0}$$

$$(iv) \quad \left. \begin{matrix} (i) \\ (iii) \end{matrix} \right\} \Rightarrow g_i = 0 : \text{contradiction}$$

c. $p_i \equiv r_i$ (so far as $r_i \neq 0$)

r_i : residual

◎ Type 1 の共役化をすれば最小化の方向は次のようになる。

a' は a の変形である。

4.3 Minimization Algorithms deduced from Type 1 Construction

a. P_1, P_2, \dots, P_n given basis

$$\begin{aligned} t_1 &= P_1 \\ t_i &= P_i - \sum_{j=1}^{i-1} \frac{P_i^T A t_j}{t_j^T A t_j} t_j \end{aligned}$$

a' P_1, P_2, \dots, P_n given basis

$$\begin{aligned} t_1 &= P_1 \\ t_i &= P_i + \sum_{j=1}^{i-1} \frac{P_i^T (g_{j+1} - g_j)}{t_j^T g_j} t_j \left(= P_i + \sum_{j=1}^{i-1} \frac{P_i^T y_j}{t_j^T g_j} t_j \right) \end{aligned}$$

$$\begin{aligned} \because A t_j &\parallel A y_j = A(x_{j+1} - x_j) \\ &= A x_{j+1} + b - (A x_j + b) \\ &= g_{j+1} - g_j = y_j \end{aligned}$$

$$\frac{P_i^T A t_j}{t_j^T A t_j} = \frac{P_i^T (g_{j+1} - g_j)}{\underbrace{t_j^T (g_{j+1} - g_j)}_0} = -\frac{P_i^T (g_{j+1} - g_j)}{t_j^T g_j} = -\frac{P_i^T y_j}{t_j^T g_j}$$

◎ この場合には，最初の基底と勾配の間に直交性がでてくる

b. given basis and gradients

$$g_{i+1}^T P_j = 0 \quad j = 1, 2, \dots, i$$

$$\because g_{i+1}^T t_j = 0 \quad j = 1, 2, \dots, i \quad (\text{Lemma 4})$$

$$0 = g_{i+1}^T \left(P_j + \sum_{k=1}^{j-1} \frac{P_j^T y_k}{t_k^T g_k} t_k \right) = g_{i+1}^T P_j$$

◎ この漸化式をさらに簡単にするために，1階の漸化式になるための条件を求める。

c. Reduction to the Recurrence Relations of the 1-st Order

$$\boxed{p_i^T A t_j = 0 \quad j=1, 2, \dots, i-2} \quad \Rightarrow \quad \boxed{t_i = p_i + \frac{p_i^T g_{i-1}}{p_{i-1}^T g_{i-1}} t_{i-1}}$$

Proof

$$\begin{aligned} t_i &= p_i - \sum_{j=1}^{i-1} \frac{p_i^T A t_j}{t_j^T A t_j} t_j = p_i - \frac{p_i^T A t_{i-1}}{t_{i-1}^T A t_{i-1}} t_{i-1} = \\ &= p_i - \frac{p_i^T (g_i - g_{i-1})}{(p_{i-1} - \frac{p_{i-1}^T A t_{i-2}}{t_{i-2}^T A t_{i-2}} t_{i-2})^T (g_i - g_{i-1})} t_{i-1} = \\ &= p_i - \frac{p_i^T (g_i - g_{i-1})}{p_{i-1}^T g_i - p_{i-1}^T g_{i-1}} t_{i-1} = \quad (\text{Lemma 4}) \\ &= p_i + \frac{p_i^T g_{i-1}}{p_{i-1}^T g_{i-1}} t_{i-1} \quad \blacksquare \end{aligned}$$

◎ 上のような条件をみたす基底として $-g_i$ が考えられる。

Fletcher Reeves の方法はこうしてみちびかれる。

Example

$p_i \equiv -g_i$ (so far as $g_i \neq 0$)

$$\begin{aligned} * \quad & p_i^T A t_j = 0 \\ & g_i^T g_j = 0 \end{aligned} \quad \left. \begin{array}{l} \\ \end{array} \right\} \quad j=1, 2, \dots, i-1$$

$$p_i^T g_j = 0 \quad j=1, 2, \dots, i-1 \Rightarrow g_i^T g_j = 0 \quad j=1, 2, \dots, i-1$$

$$\begin{aligned} p_i^T A t_j &\propto g_i^T (g_{j+1} - g_j) = +g_i^T g_{j+1} \dots = 0 \quad j=1, \dots, i-2 \\ &\quad -g_i^T g_j \dots = 0 \quad j=1, \dots, i-1 \\ &= 0 \end{aligned}$$

$$\begin{aligned} t_i &= -g_i + \frac{g_i^T (g_i - g_{i-1})}{g_{i-1}^T g_{i-1}} t_{i-1} \quad \text{--- Polak Rivière's Algorithm} \\ &= -g_i + \frac{g_i^T g_i}{g_{i-1}^T g_{i-1}} t_{i-1} \quad \text{--- Fletcher Reeves's Algorithm} \end{aligned}$$

◎ 一般に, k 階の漸化式に帰着する条件も求められる。

d. Reduction to the Recurrence Relations of the k -th Order

$$\boxed{P_i^T A t_j = 0 \quad j=1, 2, \dots, i-k-1} \Rightarrow \boxed{t_i = P_i + \sum_{j=i-k}^{i-1} \frac{P_i^T y_j}{P_j^T y_j} t_j}$$

Proof. abbr. cf. c. ■

◎ 上のような条件をみたす P_i も無くはない。(この例題の数値的な有効性はまだ調べていない。)

Example

$$P_i = -y_{i-1} = -(g_i - g_{i-1}), \quad i=2, \dots, n, \quad P_1 = -g_1$$

$$* \quad P_i^T A t_j = 0 \quad j=1, 2, \dots, i-2$$

$$\therefore P_j^T g_i = 0 \quad j=1, 2, \dots, i-1$$

$$\Rightarrow \begin{cases} g_1^T g_i = 0 \\ (g_j - g_{j-1})^T g_i = 0 \quad j=2, \dots, i-1 \end{cases}$$

$$P_i^T A t_j \propto (g_i - g_{i-1})^T (g_{j+1} - g_j) =$$

$$= g_i^T (g_{j+1} - g_j) \quad \text{----- } 0 \quad j=1, 2, \dots, i-2$$

$$- g_{i-1}^T (g_{j+1} - g_j) \quad \text{----- } 0 \quad j=1, 2, \dots, i-3$$

$$= 0 \quad j=1, 2, \dots, i-3, \quad i \geq 2$$

$$* \quad \boxed{t_i = -y_{i-1} + \frac{y_{i-1}^T y_{i-1}}{y_{i-2}^T y_{i-1}} t_{i-1} + \frac{y_{i-1}^T y_{i-2}}{y_{i-3}^T y_{i-2}} t_{i-2}}$$

◎ Type 2 の方からもっと簡単に実用的なアルゴリズムが出るか否かは未調査である。

◎ § 4 の漸化式を行列型に書き改めてみよう。

§ 5 Recurrence Relations in Matrix Form

◎ まず, Type 1 の場合については,

5.1. Recurrence Relations deduced from Type 1 Construction

a. Vector Form	Matrix Form
$\begin{aligned} \mathbf{t}_1 &= \mathbf{p}_1 \\ \mathbf{t}_i &= \mathbf{p}_i - \sum_{j=1}^{i-1} \frac{\mathbf{t}_j^T \mathbf{A} \mathbf{p}_i}{\mathbf{t}_j^T \mathbf{A} \mathbf{t}_j} \mathbf{t}_j \\ &= \left(\mathbf{p}_i + \sum_{j=1}^{i-1} \frac{\mathbf{y}_j^T \mathbf{p}_i}{\mathbf{t}_j^T \mathbf{g}_j} \mathbf{t}_j \right) \end{aligned}$	$\begin{aligned} \mathbf{t}_1 &= \mathbf{E} \mathbf{p}_1 \\ \mathbf{t}_i &= \left(\mathbf{E} - \sum_{j=1}^{i-1} \frac{\mathbf{t}_j \mathbf{t}_j^T \mathbf{A}}{\mathbf{t}_j^T \mathbf{A} \mathbf{t}_j} \right) \mathbf{p}_i \\ &= \left(\mathbf{E} + \sum_{j=1}^{i-1} \frac{\mathbf{t}_j \mathbf{y}_j^T}{\mathbf{t}_j^T \mathbf{g}_j} \right) \mathbf{p}_i \end{aligned}$
$\therefore \mathbf{t}_i = \mathbf{p}_i - \sum_{j=1}^{i-1} \frac{\mathbf{t}_j^T \mathbf{A} \mathbf{p}_i}{\mathbf{t}_j^T \mathbf{A} \mathbf{t}_j} \mathbf{t}_j = \mathbf{p}_i - \sum_{j=1}^{i-1} \mathbf{t}_j \frac{\mathbf{t}_j^T \mathbf{A} \mathbf{p}_i}{\mathbf{t}_j^T \mathbf{A} \mathbf{t}_j} = \left(\mathbf{E} - \sum_{j=1}^{i-1} \frac{\mathbf{t}_j \mathbf{t}_j^T \mathbf{A}}{\mathbf{t}_j^T \mathbf{A} \mathbf{t}_j} \right) \mathbf{p}_i$	

b. Recurrence Relations in Matrix Form

$$\mathbf{C}_i^1 \equiv \mathbf{E} - \sum_{j=1}^{i-1} \frac{\mathbf{t}_j \mathbf{t}_j^T \mathbf{A}}{\mathbf{t}_j^T \mathbf{A} \mathbf{t}_j} = \mathbf{E} + \sum_{j=1}^{i-1} \frac{\mathbf{t}_j \mathbf{y}_j^T}{\mathbf{t}_j^T \mathbf{g}_j}$$

$$\begin{aligned} \mathbf{t}_i &= \mathbf{C}_i^1 \mathbf{p}_i \\ \mathbf{C}_i^1 &= \mathbf{C}_{i-1}^1 - \frac{\mathbf{t}_{i-1} \mathbf{t}_{i-1}^T \mathbf{A}}{\mathbf{t}_{i-1}^T \mathbf{A} \mathbf{t}_{i-1}} = \left(\mathbf{C}_{i-1}^1 + \frac{\mathbf{t}_{i-1} \mathbf{y}_{i-1}^T}{\mathbf{t}_{i-1}^T \mathbf{g}_{i-1}} \right) \end{aligned}$$

Remark 1

\mathbf{C}_i^1 needs not be equal to \mathbf{E} , but it is necessary that $\mathbf{C}_i \mathbf{p}_1 \notin [\mathbf{p}_2, \mathbf{p}_3, \dots, \mathbf{p}_n]$. In this case, $\mathbf{C}_i \mathbf{p}_1$ is to be regarded as the first member of the given basis.

Remark 2

Since αC_i^1 ($i=1,2,\dots,n$ $\alpha \neq 0$) give the same search directions as C_i^1 , we may replace C_i^1 by $\tilde{C}_i^1 \equiv \alpha C_i^1$.

In this case, the recurrence relation turns out to be

$$\tilde{C}_i^1 = \tilde{C}_{i-1}^1 - \alpha \frac{\mathbf{t}_{i-1} \mathbf{t}_{i-1}^T A}{\mathbf{t}_{i-1}^T A \mathbf{t}_{i-1}}$$

Remark 3

$$C_n^1 = 0$$

see 3.4 e. Cor 3.4

Remark 4

C_i^1 ($i=1,2,\dots,n$) : idempotent

$\therefore * P_{i-1} \equiv \sum_{j=1}^{i-1} \frac{\mathbf{t}_j \mathbf{t}_j^T A}{\mathbf{t}_j^T A \mathbf{t}_j}$: conjugate projection matrix
 \downarrow
 idempotent

$$\begin{aligned} \text{cf. } \frac{\mathbf{t}_i \mathbf{t}_i^T A}{\mathbf{t}_i^T A \mathbf{t}_i} \cdot \frac{\mathbf{t}_j \mathbf{t}_j^T A}{\mathbf{t}_j^T A \mathbf{t}_j} &= \frac{\mathbf{t}_i \mathbf{t}_j^T A}{\mathbf{t}_j^T A \mathbf{t}_j} \\ \frac{\mathbf{t}_i \mathbf{t}_i^T A}{\mathbf{t}_i^T A \mathbf{t}_i} \cdot \frac{\mathbf{t}_j \mathbf{t}_j^T A}{\mathbf{t}_j^T A \mathbf{t}_j} &= 0 \quad i \neq j \end{aligned}$$

* $C_i^1 = E - P_{i-1}$: idempotent

Remark 5

$$C_j^1 \mathbf{r}_i = \mathbf{r}_i \quad j=1,\dots,i-1$$

($\mathbf{r}_i = A^{-1} \mathbf{g}_i$: residual)

$$\begin{aligned} \therefore C_{i-1}^1 \mathbf{r}_i &= C_{i-1}^1 A^{-1} \mathbf{g}_i = C_{i-2}^1 A^{-1} \mathbf{g}_i - \frac{\mathbf{t}_{i-2} \mathbf{t}_{i-2}^T A A^{-1} \mathbf{g}_i}{\mathbf{t}_{i-2}^T A \mathbf{t}_{i-2}} \\ &= C_{i-2}^1 \mathbf{r}_i = \dots = C_1^1 \mathbf{r}_i = \mathbf{r}_i \end{aligned}$$

◎ Type 2 の場合にも同様にして,

5.2 Recurrence Relation deduced from Type 2 Construction

a.	Vector Form	Matrix Form
	$\begin{aligned} \mathbf{t}_1 &= \mathbf{P}_1 \\ \mathbf{u}_1 &= \mathbf{A} \mathbf{t}_1 \\ \mathbf{u}_j &= \mathbf{A} \mathbf{t}_j - \sum_{\ell=1}^{j-1} \frac{\mathbf{u}_\ell^T \mathbf{A} \mathbf{t}_j}{\mathbf{u}_\ell^T \mathbf{u}_\ell} \mathbf{u}_\ell \\ \mathbf{t}_i &= \mathbf{P}_i - \sum_{j=1}^{i-1} \frac{\mathbf{P}_i^T \mathbf{u}_j}{\mathbf{u}_j^T \mathbf{u}_j} \mathbf{u}_j \\ i &= 1, 2, \dots, n \end{aligned}$	$\begin{aligned} \mathbf{t}_1 &= \mathbf{E} \mathbf{P}_1 \\ \mathbf{u}_1 &= \mathbf{A} \mathbf{t}_1 \\ \mathbf{u}_j &= \left(\mathbf{E} - \sum_{\ell=1}^{j-1} \frac{\mathbf{u}_\ell \mathbf{u}_\ell^T}{\mathbf{u}_\ell^T \mathbf{u}_\ell} \right) \mathbf{A} \mathbf{t}_j \\ \mathbf{t}_i &= \left(\mathbf{E} - \sum_{j=1}^{i-1} \frac{\mathbf{u}_j \mathbf{u}_j^T}{\mathbf{u}_j^T \mathbf{u}_j} \right) \mathbf{P}_i \\ i &= 1, 2, \dots, n \end{aligned}$
	$\begin{aligned} \therefore \mathbf{u}_j &= \mathbf{A} \mathbf{t}_j - \sum_{\ell=1}^{j-1} \frac{\mathbf{u}_\ell^T \mathbf{A} \mathbf{t}_j}{\mathbf{u}_\ell^T \mathbf{u}_\ell} \mathbf{u}_\ell \\ &= \mathbf{A} \mathbf{t}_j - \sum_{\ell=1}^{j-1} \mathbf{u}_\ell \frac{\mathbf{u}_\ell^T \mathbf{A} \mathbf{t}_j}{\mathbf{u}_\ell^T \mathbf{u}_\ell} \\ &= \left(\mathbf{E} - \sum_{\ell=1}^{j-1} \frac{\mathbf{u}_\ell \mathbf{u}_\ell^T}{\mathbf{u}_\ell^T \mathbf{u}_\ell} \right) \mathbf{A} \mathbf{t}_j \end{aligned}$	$\begin{aligned} \mathbf{t}_i &= \mathbf{P}_i - \sum_{j=1}^{i-1} \frac{\mathbf{P}_i^T \mathbf{u}_j}{\mathbf{u}_j^T \mathbf{u}_j} \mathbf{u}_j \\ &= \mathbf{P}_i - \sum_{j=1}^{i-1} \mathbf{u}_j \frac{\mathbf{u}_j^T \mathbf{P}_i}{\mathbf{u}_j^T \mathbf{u}_j} \\ &= \left(\mathbf{E} - \sum_{j=1}^{i-1} \frac{\mathbf{u}_j \mathbf{u}_j^T}{\mathbf{u}_j^T \mathbf{u}_j} \right) \mathbf{P}_i \end{aligned}$

b. Recurrence Relation in Matrix Form

$$\mathbf{C}_i^2 \equiv \left(\mathbf{E} - \sum_{j=1}^{i-1} \frac{\mathbf{u}_j \mathbf{u}_j^T}{\mathbf{u}_j^T \mathbf{u}_j} \right)$$

$$\begin{aligned} \mathbf{t}_1 &= \mathbf{C}_1^2 \mathbf{P}_1 \\ \mathbf{C}_i^2 &= \mathbf{C}_{i-1}^2 - \frac{\mathbf{C}_{i-1}^2 \mathbf{y}_{i-1} \mathbf{y}_{i-1}^T \mathbf{C}_{i-1}^{2T}}{(\mathbf{C}_{i-1}^2 \mathbf{y}_{i-1})^T (\mathbf{C}_{i-1}^2 \mathbf{y}_{i-1})} \end{aligned}$$

$$\begin{aligned} \therefore \mathbf{C}_i^2 &= \mathbf{C}_{i-1}^2 - \frac{\mathbf{u}_{i-1} \mathbf{u}_{i-1}^T}{\mathbf{u}_{i-1}^T \mathbf{u}_{i-1}}, \quad \mathbf{u}_{i-1} = \mathbf{C}_{i-1}^2 \mathbf{A} \mathbf{t}_{i-1} \\ &= \mathbf{C}_{i-1}^2 - \frac{\mathbf{C}_{i-1}^2 \mathbf{A} \mathbf{t}_{i-1} \mathbf{t}_{i-1}^T \mathbf{A} \mathbf{C}_{i-1}^{2T}}{\mathbf{t}_{i-1}^T \mathbf{A} \mathbf{C}_{i-1}^{2T} \mathbf{C}_{i-1}^2 \mathbf{A} \mathbf{t}_{i-1}} \\ &= \mathbf{C}_{i-1}^2 - \frac{\mathbf{C}_{i-1}^2 \mathbf{y}_{i-1} \mathbf{y}_{i-1}^T \mathbf{C}_{i-1}^{2T}}{(\mathbf{C}_{i-1}^2 \mathbf{y}_{i-1})^T (\mathbf{C}_{i-1}^2 \mathbf{y}_{i-1})} \end{aligned}$$

Remark 1

C_i^2 needs not be equal to E . cf. Remark 1 in 5.1, b.

Remark 2

C_i^2 may be replaced by $\tilde{C}_i^2 \equiv \alpha C_i^2$, which is given by the recurrence relation

$$\tilde{C}_i^2 = \tilde{C}_{i-1}^2 - \alpha \frac{C_{i-1}^2 y_{i-1} y_{i-1}^T C_{i-1}^2}{(C_{i-1} y_{i-1})^T (C_{i-1} y_{i-1})}$$

cf. Remark 2 in 5.1 b

Remark 3

$$C_n^2 = 0$$

see 3.4 e. Cor. 3.4

Remark 4

C_i^2 : symmetric, idempotent

(orthogonal projection to $[At_1, \dots, At_{i-1}] = [y_1, \dots, y_{i-1}]$)

Remark 5

$$C_i^2 y_l = y_l \quad l \leq i-1$$

$$\therefore C_i^2 y_l = \left(E - \sum_{j=1}^{i-1} \frac{u_j u_j^T}{u_j^T u_j} \right) y_l$$

$$= y_l \quad u_j \in [At_1, \dots, At_{i-1}] \perp t_i, t_{i+1}, \dots$$

- ◎ 最小点と共に, A^{-1} も求めたい場合には,

5.3 Construction of A^{-1}

$$* A^{-1} = \sum_{i=1}^n \frac{\mathbf{t}_i \mathbf{t}_i^T}{\mathbf{t}_i^T A \mathbf{t}_i} \quad (3.4, d. Cor. 3.3)$$

$$= \sum_{i=1}^n \frac{\mathbf{w}_i \mathbf{w}_i^T}{\mathbf{w}_i^T A \mathbf{w}_i} \quad \because \mathbf{t}_i \parallel \mathbf{w}_i$$

$$* D_i \equiv \sum_{j=1}^{i-1} \frac{\mathbf{t}_j \mathbf{t}_j^T}{\mathbf{t}_j^T A \mathbf{t}_j} = \sum_{j=1}^{i-1} \frac{\mathbf{w}_j \mathbf{w}_j^T}{\mathbf{w}_j^T A \mathbf{w}_j} \Rightarrow D_{nn} = A^{-1}$$

$$\boxed{\begin{aligned} D_1 &= 0 \\ D_i &= D_{i-1} + \frac{\mathbf{t}_{i-1} \mathbf{t}_{i-1}^T}{\mathbf{t}_{i-1}^T A \mathbf{t}_{i-1}} \end{aligned}} \Rightarrow D_{nn} = A^{-1}$$

$$* D_i : \text{symmetric}$$

$$* D_i \mathbf{g}_k = 0 \quad k \geq i$$

$$\because D_i \mathbf{g}_k = \sum_{j=1}^{i-1} \frac{\mathbf{w}_j \mathbf{w}_j^T \mathbf{g}_k}{\mathbf{w}_j^T A \mathbf{w}_j} = 0$$

①さらに, A^{-1} をもとめる漸化式と最小値を求める漸化式をまとめ、1つの漸化式を作ることできる。

5.4 Combined Algorithms

Example (Combined Algorithm, Davidon-Fletcher-Powell)

$$\begin{aligned} P_i &= g_i \\ H_i^2 &= C_i^2 + D_i \\ t_i &= H_i^2 P_i \end{aligned}$$

$$* \quad H_i^2 g_k = C_i^2 g_k \quad k \geq i,$$

$$\begin{aligned} \therefore H_i g_i &= (C_i^2 + D_i) g_i \\ &= C_i^2 g_i + \sum_{j=1}^{i-1} \frac{t_j t_j^T}{t_j^T A t_j} g_i \\ &= C_i^2 g_i \end{aligned}$$

$$* \quad H_i^2 y_k = C_i^2 y_k \quad k \geq i$$

$$* \quad H_i^2 \text{ symmetric}$$

$$H_i^2 = H_{i-1}^2 - \frac{H_{i-1}^2 y_{i-1} y_{i-1}^T H_{i-1}^2}{(H_{i-1}^2 y_{i-1})^T (H_{i-1}^2 y_{i-1})} + \frac{v_{i-1} v_{i-1}^T}{v_{i-1}^T A v_{i-1}}$$

Example (Combined Algorithm)

$$\begin{aligned} P_i &= g_i \\ H_i^1 &= C_i^1 + D_i \\ t_i &= H_i^1 P_i \end{aligned}$$